TREND LINES AND THE NUMBER OF SPECIES OF STAPHYLINIDAE

J. H. FRANK AND G. A. CURTIS

Florida Medical Entomology Laboratory, P.O. Box 520,
Vero Beach, FL 32960; Indian River Mosquito Control District,
P.O. Box 670, Vero Beach, FL 32960

ABSTRACT

The limitations of the use of trend lines to predict the number of species existing are explained. The Cicindelidae of America north of Mexico are used as an example to explain the fitting of calculated trend lines as opposed to the rule-of-thumb methods which have been used by previous authors. The limitations of interpretation of calculated trend lines are explained. The number of species of Staphylinidae of America north of Mexico is estimated as > 3,416 by a simple method independent of trend lines and it is demonstrated that the use of trend lines gives an erroneous prediction of this total. The number of species of Staphylinidae of the world cannot at present be estimated with any accuracy because of paucity of data suitable for analysis.

INTRODUCTION

Trend lines, of the cumulative number of species described vs. time, have been used in attempts to predict the number of species existing in various taxa by Steyskal (1965), Arnett (1967) and White (1975). Our initial examination of these publications was marked with a certain amount of incredulity as to methods used and assumptions made.

Our routine work requires the fitting of calculated regression lines to biological data, and trend lines are regression lines. One of us has access to the literature on Staphylinidae, so we decided to compare several types of regression lines using figures for this group of the Coleoptera. We decided also, lest the figures for Staphylinidae are in some way exceptional, to compare regression lines in 2 other families of Coleoptera, and for these other examples we selected the Cicindelidae of America north of Mexico, and the Curculionidae. The Cicindelidae were selected because numbers of species described and a trend line had been presented by White (1975) and the trend line illustrated appears to have reached an upper asymptote. The Curculionidae were selected because they are known to be a very large family, and if regression analysis should for some unforeseen reason differ in large families (e.g. Staphylinidae) from that in small families (e.g. Cicindelidae), then analyses of curves for Staphylinidae and Curculionidae might act as useful cross-references to one another.

Data for Cicindelidae were derived from White (1975), while data for Curculionidae were supplied by C. W. O'Brien. Explanation of the fitting of calculated regression lines is made in this article, using Cicindelidae as an example. Analysis is also made here of the number of species of Staphylinidae. Analyses for Curculionidae have been completed and are presented separately by O'Brien & Wibmer.

Estimates of the total number of species of Staphylinidae have been made previously. For America north of Mexico, Arnett (1967) estimated 3,500. The following estimates are for the world total. Blatchley (1910) stated: "Sharp says that it is probable that one-hundred thousand species or even more of Staphylinidae are at present in existence". Edwards (1949) cited the same figure of 100,000. Arnett (1967) suggested a considerably lower total of 28,000, although Seevers (1965) believed more than 28,000 species already had been described. More recently, Hammond (1975) has indicated that the subfamily Aleocharinae alone may contain more than 100,000 species. These estimates, without any supportive explanation, are evidently highly speculative but do suggest that the family is larger than are most other families of Coleoptera. Fowler (1888) made a speculative claim that: "the family Staphylinidae probably contains more species than any other family of Coleoptera".

Basic Assumptions

The very first tenet which should be examined is the reason for using trend lines for estimation of the total number of species existing. What we wish to estimate is the total number of species in a taxon. It seems to us unnecessary to attempt also to estimate the year in which all of these species will have been described. Perhaps an estimate of the year is seen as an additional benefit of the use of the method, but in trying to estimate both the year and the number we are adding greatly to the complexity. We suspect that although some insect taxonomists would be willing to hazard a guess at the total number of species within taxa known to them, fewer would risk guessing the year in which all species will have been described and would possibly answer that the date would depend entirely upon the amount of time devoted to the task. In this article we have used the method of trend line fitting because, in one form or another, it has been used before, thus some evidently believe it to be a valid method, but the purpose of this article is as much to evaluate the method as to derive estimates by its use.

The trend lines illustrated by White (1975) for some families of Coleoptera of America north of Mexico are sigmoidal in form. They indicate, some more clearly than others, that for the first 50 to 100 years since Linné (1758), the number of species described from the region was relatively small. That is, the number of species described during each 10 year period was small and the cumulative numbers (plotted in the graphs) show only a slight gradient. A marked change is apparent in the middle third of the 19th century, when the trend lines show an increased slope. This upturn represents an increased output of published species descriptions by entomologists, in other words: the effort put into the collection of specimens and publications of species descriptions, called here for want of a better term DESCRIPTIVE EFFORT, increased. The reasons for this increased effort are only of historical interest. The trend lines show a nearly or entirely linear climb until after the turn of the present century, when most of them show signs of levelling off. Unlike the initial upturn, which we hold to be of no more than historical interest, this levelling off must be considered carefully if we are to attempt to make any predictions as to the future slope of the curve.

Contrary to the suggestions by White (1975), there are no *a priori* reasons why the curve marked by the levelling off should be a mirror image of the initial upturn. If there has been a reduction in DESCRIPTIVE EFFORT, then there is no reason to imagine its rate to have been in precisely inverse direction to that causing the upturn. If the upper curve should be due largely or entirely to a reduction in DESCRIPTIVE EFFORT, then the trend line is of no value as an indicator of the total number of species existing. However, we know of no historical reasons to believe that there has been a massive reduction in DESCRIPTIVE EFFORT and, in order to make any sense of this study, we are forced to assume (ASSUMPTION NO. 1) that: any change in DESCRIPTIVE EFFORT since the middle third of the 19th century is negligible.

In contrast to the above, it must be assumed (ASSUMPTION NO. 2) that: the levelling off of the curve is the result entirely of the description of new species becoming increasingly difficult because of the decreasing probability of discovery. In other words, as the description of species approaches totality, so it becomes increasingly uncommon for an undescribed species to be discovered, despite undiminished DESCRIPTIVE EFFORT.

Unfortunately, the 2 assumptions are not only tenuous but also impossible to evaluate. Although the function of the taxonomist always has been nominally to classify organisms, the reader of taxonomic publications of the 19th and even early part of the 20th century might suspect that the primary emphasis of many of the authors was to write species descriptions, perhaps even that some of the authors attached some merit to the number of their published species descriptions. Most, if not all, modern taxonomists would deny emphatically any especial merit in publication of a large number of species descriptions and would instead stress the importance of classification. Thus, modern species descriptions are incidental to the function of the taxonomist, and are written with the intention of producing adequate tools for the classification of species within genera, not merely of species recognition and allocation to genus.

Because species descriptions published in the 19th and early part of the 20th century often fail to provide adequate tools for classification of species within genera, the modern taxonomist may be obliged to rewrite them, and this inevitably reduces the effort which can be devoted to describing previously unrecognised species. His publications are often in the form of thorough revisions of genera or supra-generic taxa, which may demand examination of large numbers of specimens borrowed from many collections. All of this allows at least the possibility that DESCRIPTIVE EFFORT (as defined above) has indeed been reduced.

The work of the modern taxonomist is in some ways easier than that of his predecessors, because of modern technological advantages: better equipment, communications, availability of literature and type material, faster travel and even (for a few) the services of technicians, typists, photographers, illustrators, translation services and computerized informative storage and retrieval. Variation in numbers of taxonomists and in their individual and collective DESCRIPTIVE EFFORTS further compounds the difficulty of evaluating the assumptions. This maze of variables with possible effect upon DESCRIPTIVE EFFORT makes acceptance of the assumptions a faith instead of an exercise in statistics. If either assumption is demonstrably false, then it is pointless to attempt to make predictions from a trend line.

Then, we must assume (ASSUMPTION NO. 3) that no species have evolved, nor have any become extinct, since the time of Linné, nor will any become extinct, during the future which we attempt to predict. The concept that natural or man-made disasters might cause species extinction cannot be taken into consideration.

Next, there is the difficulty caused by species synonymies. Our data should be the number of species described, free of undiscovered synonyms, for every data point. Yet this can be possible only in a few families, those in which there are no undiscovered synonyms and where each synonym may be discounted from the total number of species known back to the year in which the synonymy was inadvertently caused. An up-to-date catalogue or index which is believed to contain no undiscovered synonyms would give this information in appropriate form for direct use. If the catalogue is expected to contain undiscovered synonymies then there are 2 options, neither of which is very good: (1) the existence of undiscovered synonyms may be ignored, or (2) the number of species listed in the latest catalogue may be taken as the latest data point, while old catalogues may be consulted for the number of species believed to have been described up until each of the appropriate dates of catalogue publication. In some families, with a large percentage of undiscovered synonyms, trend curves constructed may lead to entirely erroneous conclusions. White (1975) has hinted at this difficulty with regard to the taxonomic work by T. L. Casey on certain families of Coleoptera of the United States, yet Casey was far from the only worker to cause synonymies.

We have not yet discussed all of the necessary assumptions or difficulties relative to trend curve analysis, but the remainder are easier to explain by reference to the actual examples which follow.

The Fitting of Regression Lines for Cicindelidae

Examining Fig. 4 in the article by White (1975) we completed columns x and y of Table 1, where the x values indicate the dates 1770, 1780 . . . 1970 at equispaced intervals of time, and the y values indicate the cumulative number of species described up until each of the dates. It is probable that we have made errors in estimating the appropriate y values from the graph (Fig. 4) but we have no doubt that any such errors are entirely negligible.

The distribution of data points in the graph indicates clearly, before the fitting of regression lines, that a sigmoidal relationship exists, therefore we should use a regression equation which is able to give a sigmoidal line. The fitting of a calculated sigmoidal regression line differs in several respects from the crude method propounded by White (1975): (1) no assumption is made that the upper part of the curve will be an exact match in mirror image to the lower part; (2) no assumption is made that the line must pass through the first and last data points; (3) no assumption is made that we can guess the mid-point of the line segment with accuracy—it is, of course, illogical to imagine that the mid-point of any line can be known until both end-points are known.

Firstly we simplified the x values in Table 1 by subtracting 1769 from each of them, to give the column headed x_1. This makes no difference to the outcome of the calculation, but saved a certain amount of button-pushing on the keyboard of a calculator. Then we calculated curves using 3 types of regression equation, all giving sigmoids: (1) cubic, (2) log quadratic, and (3) logistic, using a programmable calculator. Anyone unfamiliar with calculation of these regressions may refer to a textbook such as Bliss (1967,

Table 1. Data points, estimates and extrapolations for Cicindelidae of America north of Mexico

x	x_1	y	$\hat{y}_c$	$\hat{y}_{log\ q}$	$\hat{y}_{logistic}$	1930
1770	1	2	3	3	3	3
1780	11	5	3	4	4	4
1790	21	8	5	6	6	6
1800	31	8	8	9	8	9
1810	41	9	13	13	12	12
1820	51	15	20	18	17	17
1830	61	28	27	24	24	23
1840	71	42	35	32	32	31
1850	81	46	44	41	42	41
1860	91	55	54	51	53	51
1870	101	61	63	62	65	63
1880	111	73	73	74	76	74
1890	121	80	82	86	87	85
1900	131	85	91	96	95	94
1910	141	100	98	106	102	101
1920	151	110	105	113	108	107
1930	161	112	111	118	112	112
1940	171	115	116	120	115	115
1950	181	117	118	118	117	117
1960	191	118	119	114	118	119
1970	201	119	118	107	119	120
1980	211	-	114	98	120	121
1990	221	-	108	87	120	121
2000	231	-	100	75	120	122
2010	241	-	88	63	121	122
2020	251	-	73	52	121	123

x = year, x_1 = year - 1769, y = actual no. of species recorded, $\hat{y}_c$ = estimates by cubic method, $\hat{y}_{log\ q}$ = estimates by log quadratic method, $\hat{y}_{logistic}$ = estimates by logistic method, 1930 = estimates by logistic method when only data to 1930 are used to derive estimates. Figures below the line are extrapolations.

1970) for detailed explanation. The estimated values for y ($\hat{y}$ signifies an estimated value as opposed to y, the actual data point) by each of the 3 methods are shown in Table 1.

The cubic regression estimates are made by the formula:

$\hat{y}_c = a' + c_1x_1 + c_2x_1^2 + c_3x_1^3$, the constants $a' = 3.5579$, $c_1 = -1.5399 \times 10^{-1}$, $c_2 = 1.1185 \times 10^{-2}$, $c_3 = -3.7746 \times 10^{-5}$ having been determined by solving a set of algebraic equations.

The log quadratic estimates are made by the formula:

$y_{\log q} = 10^{x}\ (a' = q_1\ x_1 + q_2\ x_1{}^2)$, where 10^{x} = antilogarithm to the base 10, the constants $a' = 0.3943$, $q_1 = 1.9636 \times 10^{-2}$, $q_2 = -5.7249 \times 10^{-5}$ having been determined by solving algebraic equations.

The logistic estimates are made by the formula:

$$\hat{y}_{\text{logistic}} = \frac{c}{100}\left(100 - \left(\frac{100}{1 + \exp\left(a' + b\left(\left(\ln \frac{100x_1}{c}\right) - \left(\ln\left(100 - \frac{100}{x_1}\right)\right)\right)\right)}\right)\right),$$

where 1n = logarithm to the base e, exp = antilogarithm to the base e, the constants $a' = -3.8012$ and $b = 0.0390$ having been determined by solving an equation, the constant $c = 121$, i.e. the estimated upper asymptote, having been determined by an iterative method involving successive approximations until the best fit was obtained by a least-sum-of-squares method using the transformed data.

Each value of the 3 sets of estimates is rounded off to the nearest whole number in Table 1. Each set will produce a smooth curve when graphed. The question now arises as to which of the 3 sets of estimates is to be preferred and what should be the basis of the selection.

Normal use of fitted regression lines requires only that estimates be made within the limits of the data, e.g. with the data as presented in Table 1 we may make reasonable estimates of the number of described species for any year between 1770 and 1970, but neither before nor after this 200-year time period. Extrapolation, that is the prediction of $\hat{y}$ values beyond the period for which we have data (i.e. here the 200-year period) is, at best, tentative, and the further into the future the extrapolation is made, the less likely it is to be accurate. Note that we can be as precise as we like since we can take the estimates (rounded off in Table 1) to as many places beyond the decimal point as we like, yet precision beyond the decimal point is evidently meaningless because a species can only be represented as a whole number. Precision on the "whole number side" of the decimal point is also something to be wary of since the estimates do differ, if slightly, from the data points. Within the 200 year period, we determined that the cubic equation provides the best empirical fit to the data points because the sum of the squares of the deviation of the estimates from the data points, i.e. $\sum (y-\hat{y})^2$, is the least, having a value of 188, the logistic equation provides the next best fit, with a value of 337, and the log quadratic the worst, with a value of 615. Thus, for normal biological purposes we would probably select the cubic equation.

In an attempt to predict the course of events in the future, we have extrapolated estimates beyond the present time in Table 1. By the cubic equation, it is evident that the trend line has reached its upper limit at 119 (against the year 1960) and is beginning to decline, in fact, will decline indefinitely. The log quadratic equation reached its upper limit of 120 (against the year 1940) and is declining; it, too, will decline indefinitely. The logistic curve continues to rise to an upper asymptote of 121, but the increments as it approaches 121 become progressively minute although it has exceeded

120.5 (shown in Table 1 as 121) by the year 2010. The three curves, calculated independently from the same data, suggest that either all species are known or at most 2 more species (to give a maximum of 121) remain to be recognised in North America. This estimate depends upon whether the assumptions (explained earlier) are justified for Cicindelidae and whether the data are accurate. We cannot pretend that 120.5 species will be described by the year 2010, nor that declining numbers following the attainment of the upper limit (suggested by the cubic and log quadratic regressions) have any meaning.

As an example of what would have been predicted had we attempted to make the prediction in the 1930's we recalculated the logistic curve ignoring the last 4 data points. The estimates are given in Table 1 in the column headed 1930. It is evident that many of these estimates differ slightly from the estimates made by using all the data. The upper asymptote estimate of 123 species gave the best fit. This shows very clearly that the number of yet-undescribed species, together with the date of their eventual description, can effect the entire course of the calculated trend line. This alone is destructive of any argument in support of the predictive value of trend lines.

We have fitted 3 types of regression, all of them capable of giving a sigmoidal line, to the data. We would be inclined to use the cubic estimates were we not obliged to make extrapolations. However, extrapolations made to dates earlier than 1758 and later than 1970 eventually provide totally unrealistic estimates, whether of infinitely small or infinitely large numbers. Since we are obliged to assume that the number of species described was zero prior to Linné (1758) and that there is a fixed upper limit to the number of existing species, then the regression equation we use must provide both a lower and an upper asymptote. Thus, we must reject the cubic and log quadratic equations, even if they provide a better fit to the data, and use the logistic equation because only it of the 3 provides the asymptotes.

Examining the graph for Cicindelidae provided by White (1975), it is apparent that several of the data points fall on one side or other of the trend line. This is even more evident in the graph for Hydrophilidae (*ibid.*). Runs of data points on one side or another of the line would occur also were we to use a fitted logistic trend line. The distribution of these points is clear evidence that DESCRIPTIVE EFFORT was not even, but that more effort was made during certain decades, or runs of successive decades, than in others. We discuss the reason for this when we deal with the Staphylinidae of America north of Mexico. Meanwhile, we point out that this scatter of points (a) prevents an optimal fit of the trend line, (b) that its occurrence is more clearly discerned when we use non-cumulative numbers (e.g. Table 3, column y), (c) that its occurrence is neither regular nor completely random, but represents a sort of shotgun effect, (d) that while in some cases it may not completely invalidate our assumption no. 1, it reduces the accuracy of predictions made by extrapolating the trend lines, and (e) that were we to modify our logistic regression equation to take account of it we would not only be forced to use a much more complex equation, but extrapolations made by using such a complex equation would be no more accurate than those made by the logistic equation we have explained.

In the section headed BASIC ASSUMPTIONS we stated that the initial upturn of the trend line was due to increased effort and that the reasons for the increased effort are only of historical interest. There would thus be some justification for ignoring all of the earlier data points and using only the data points later than some point in the mid- or late 19th century for the calculation. This would have the advantage that a regression equation giving a single (upper) asymptote could be used, and the calculation would be simplified. However, the selection of a "starting point" would be arbitrary and different "starting points" would produce different estimated trend lines because of imperfect linearity of the data.

The problems involved in making accurate predictions from trend lines approach a magnitude where other methods of making estimates are unquestionably to be preferred. Two methods occur to us. One of these would demand initiation of intensive systematic collections from designated areas of entire faunal regions. The material collected would be identified as far as possible and the ratio of undescribed to described species represented in these collections would be apportioned to the known number of species from the entire region. This, however, would be totally impracticable merely for the present purposes for several reasons, and additionally would be subject to sampling error. We have used the second method in making an estimate of the number of species of Staphylinidae of America north of Mexico. Its extreme simplicity makes it the method of choice wherever it can be used, but its applicability depends upon the nature of recent taxonomic publications concerning a given taxon of a given faunal region.

Staphylinidae of America North of Mexico

A conspectus of recent taxonomic revisions of the group gives some pertinent information, as shown in Table 2. The number of species dealt with in the revisions listed was 320 (Table 2, column B), of which 99 (Table 2, column C) were described as new, i.e. 31%. Evidently the staphylinid fauna of the region is far from completely known. To add to the figure of 99 newly described species, the presence of 4 introduced species (Table 2, column D) was recorded for the first time and 3 species names (Table 2, column E) were removed from synonymy, so that it may be stated that ((99 + 4 + 3) × 100 ÷ 316 = 34%) or a minimum of one third of the species of the region are as yet unrecognized. We state deliberately a minimum of one third because we have reason to believe that not all of the species of the groups revised have yet been described.

This fraction of one third is, however, deceptive. We find (Table 2, column F) that 66 species names were newly placed in synonymy and that (Table 2, column G) the presence of 1 (palearctic) species in North America is doubted. Thus the number of species recognized in the groups was 281 before revision (Table 2, column A) and 320 after revision (Table 2, column B), representing a lesser increase than would have been expected by considering only the statement that a minimum of one third of the species are as yet unrecognized.

Summing the apparently valid species names as listed in the catalogue by Moore & Legner (1975) and excluding the family Micropeplidae (elsewhere included as the subfamily Micropeplinae of the Staphylinidae) we find that approximately 3,000 species were recognized in 1970 (Table 3). Ignoring any discrepancy between the total known in 1968 (we have used this date as cut-off point in Table 2) and 1970, then a minimum of 3,000 ×

Table 2. New species and synonymies in recent (1968-1977) taxonomic revisions of Staphylinidae of America north of Mexico.

Taxon	Publication	A	B	C	D	E	F	G
Bledius (part) & related genera	Herman 1972b, 1976	68	51	6	0	0	13	0
Charhyphus	Herman 1972a	1	2	1	0	0	0	0
Coproporus & *Cilea*	Campbell 1975b	11	10	0	0	0	1	0
Erichsonius	Frank 1975	9	17	10	0	0	1	1
Goniusa	Kistner 1976	1	2	1	0	0	0	0
Oxyporus	Campbell 1969	12	10	0	0	0	2	0
Pseudopsinae	Herman 1975	3	8	6	0	0	1	0
Quediini	Smetana 1971a, b, 1973, 1976	102	129	49	0	0	22	0
Sepedophilus	Campbell 1976	22	30	13	1	2	8	0
Stilicolina	Herman 1970	4	5	1	0	0	0	0
Tachinomorphus	Campbell 1973b	2	4	2	0	0	0	0
Tachinus	Campbell 1973a, 1975 Ulrich & Campbell 1974	38	43	9	3	0	7	0
Xenodusa	Hoebeke 1976	5	5	0	0	1	1	0
Zalobius, *Asemobius*, *Nanobius*	Herman, 1977	3	4	1	0	0	0	0
		281	320	99	4	3	66	1
					106		67	

A = no. of spp. recognized before revision; B = no. of spp. recognized after revision; C = no. of new spp. described in revision; D = no. of introduced spp. first recorded in revision; E = no. of spp. removed from synonymy in revision; F = no. of spp. placed in new synonymy in revision; G = no. of spp. whose presence in the region is doubted as result of revision.

320 ÷ 281 = 3,416 species should exist in America north of Mexico. The only assumption we have had to make is that the taxa revised recently (Table 2, column B) with 320 species give a sample which is representative of the 3,000 or so recognised species. We were able to use this simple method because of the status of taxonomic work on the Staphylinidae of America north of Mexico. Before the publication of the revisions listed in Table 2, it could fairly be stated that practically every genus of the family as represented in the faunal region needed revision; therefore, we believe that

the taxa listed in Table 2 were not selected for revision because they were thought to be specially in need of revision, but that they form a reasonably random sample. Although the poorly-known subfamily Aleocharinae is under-represented, and the estimate of more than 3,416 may thus be somewhat low, we have a sample size of better than 10%. There are statistical methods available for determining sample size necessary to make predictions with various levels of accuracy, but we do not have the option of increasing our sample size should this be necessary. When more revisions are completed, so as to give another sample of better than 10%, we shall be able to check, and adjust if necessary, the estimate made here.

Thus, based on only one assumption, we have reason to believe that more than 3,416 species of Staphylinidae occur in America north of Mexico. Quite how many more than 3,416 species there might be, we cannot say. However, the generic revisions listed in Table 2 used as material not merely specimens collected by the various authors, but most or all specimens available from most or all major collections having a significant amount of North American material, so that the figure of 3,416 is unlikely to be a gross under-estimation. Probably, nearly all yet-undescribed species of this family occurring in this region are represented by specimens in some collection.

We prepared the x and y columns of Table 3 from figures obtained from the catalogue by Moore & Legner (1975). Columns x and x_1 give dates much as in Table 1, column y gives non-cumulative numbers of species, column y_1 gives cumulative numbers of species as in Table 1. Probably, we have made errors in recording the y column, but we have no doubt that these errors are negligible. We note that the authors of the catalogue have included information for 1973 and in some cases for 1974 and that during this first third of the decade of the 1970's about 82 species were described, but we have not included this figure in Table 1. The column headed y shows, for some decades, several figures higher than those for earlier and later decades. Thus, the figure of 115 for 1810 is high (due largely to the work of Gravenhorst 1806), likewise 188 for 1840 (due largely to the work of Erichson 1839-40), likewise the figures for some but not all of the decades from 1890 to 1920 (due largely to the work of T. L. Casey), and for 1960 (due to the work of M. H. Hatch). These exceptional decades indicate that descriptive effort was not even, thus assumption no. 1 (explained earlier) is not well-justified and the fitting of a good logistic regression line to the y_1 data, i.e. a line where estimates match actual values closely, will not be possible. Having also discovered the high percentage of synonymy occurring in the literature (Table 2), we expect that this too will cause difficulty in the fitting of a trend line and are thus warned that the effort involved in attempting to fit a line will almost certainly be wasted. To show that such a line will demonstrably be erroneous, we have estimated an upper asymptote from the data given in Table 3.

Calculating the line of best-fit using the logistic method, we find that estimated upper asymptotes of 3,500, 3,400, 3,300, 3,200 and 3,100 give progressively better fits to the data, thus the estimated total is less than 3,100. We cannot calculate a line for 3,000 or less using the logistic formula because the equation demands that no data point exceed the estimated asymptote, so we cannot state that an asymptote of 3,000 or 2,900 would lead to a better fit. Clearly, however, the estimate of $<3{,}100$ is considerably lower

Table 3. Data points for Staphylinidae of America north of Mexico.

x	x_1	y	y_1
1760	1	8	8
1770	11	4	12
1780	21	9	21
1790	31	7	28
1800	41	13	41
1810	51	115	156
1820	61	3	159
1830	71	17	173
1840	81	188	361
1850	91	48	409
1860	101	110	519
1870	111	145	664
1880	121	189	853
1890	131	394	1,247
1900	141	146	1,393
1910	151	871	2,264
1920	161	374	2,638
1930	171	69	2,707
1940	181	48	2,755
1950	191	21	2,776
1960	201	197	2,973
1970	211	30	3,003

The figures under column y = non-cumulative no. of species, y_1 = cumulative no. of species.

than our independent estimate of > 3,416, and we have more reason to accept the independent estimate because in calculating it we have not knowingly violated any basic assumptions. The estimate derived by the fitting of a trend line is clearly erroneous.

Staphylinidae of the World

We are able to make an independent estimate of the number of staphylinid species of America north of Mexico by examining recent taxonomic revisions. Unfortunately, revisions of taxa at the generic or higher levels are seldom made for the entire world, being more frequently restricted to a faunal region. We know of only 2 recent revisions on a world basis.

Herman (1975) has revised the Pseudopsinae of the world and has found that 24 of the 30 recognized species were previously undescribed, i.e. 24 × 100 ÷ 30 = 80% of the species were not known previously. Campbell (1973a, 1975), Ulrich & Campbell (1974) and Ulrich (1975) have revised the genus *Tachinus* and described 51 new species out of a total of 158 recognized, that is 51 × 100 ÷ 158 = 31% of the species were not known previously. The Pseudopsinae do have a worldwide distribution but seem to be restricted to montane areas. The genus *Tachinus* is largely holarctic in distribution and the insect fauna of the holarctic region is better known than that of other areas, thus it would not be expected that *Tachinus* would contain a high percentage of undescribed species. It is instructive to discover that the *Tachinus* subgenus *Tachinoplesius,* with an afrotropical (Crosskey & White 1977) distribution now has 7 recognized species while before Ulrich's (1975) revision it contained only 2, thus 5 × 100 ÷ 7 = 71% of the species were found to be undescribed; it is also probable that there are as yet undescribed species of *Tachinoplesius.* As further evidence of the high percentage of undescribed species in the afrotropical fauna, Fagel's (1970) revision of some of the genera of Pinophilini in that region indicated 161 previously undescribed species out of a total of 205, i.e. 161 × 100 ÷ 205 = 79% of undescribed species. It is likely that the neotropical staphylinid fauna is about as poorly known as is the afrotropical, with the australasian and oriental perhaps somewhat better-known. These few publications do not provide a large enough sample for an independent estimate; all that we can say is that there is probably a much larger percentage of undescribed species in the world fauna than in the holarctic or nearctic faunas.

We shall attempt to fit regression lines to data for the world fauna, but we suspect that little confidence may be placed in estimates so made. To do this we completed the x and y columns of Table 4, having obtained the data from published estimates and catalogues as specified in the following paragraph.

The data points are derived from the following publications: 1758 (Linné 1758 total), 1775 (Fabricius 1775 total), 1787 (Fabricius 1787 total), 1792 (Fabricius 1792 total), 1798 (Fabricius 1798 total), 1801 (Fabricius 1801 total), 1806 (Gravenhorst 1806 total), 1831 (Mannerheim 1831 total), 1840 (Erichson 1839-40 total), 1868 (Gemminger & Harold 1868, fide Ganglbauer 1895: 15), 1872 (Fauvel 1872: 4), 1883 (Duvivier 1883, fide Ganglbauer 1895: 15), 1934 (Bernhauer *et al.* 1910-1926 + Scheerpeltz 1933-34, fide Arnett 1961: 235), 1957 (Seevers 1957: 60), 1965 (Seevers 1965: 141). The total number of species listed in both parts of the Coleopterorum Catalogus is given as the

Table 4. Data points, estimates and extrapolations for Staphylinidae of the world.

x	x_1	y	$\hat{y}_c$	$\hat{y}_{\log q}$	$\hat{y}_{logistic}$
1758	1	19	581	11	17
1775	18	25	2	32	38
1787	30	45	-187	68	69
1792	35	89	-208	91	87
1798	41	95	-185	129	117
1801	44	107	-154	153	135
1806	49	394	-72	201	172
1831	74	402	939	717	570
1840	83	1,507	1,565	1,083	874
1868	111	4,000	4,492	3,367	3,134
1872	115	4,500	5,038	3,887	3,721
1883	126	6,500	6,716	5,630	5,830
1934	177	19,909	18,163	19,803	22,030
1957	200	25,000	25,502	27,260	26,593
1965	208	28,000	28,401	29,386	27,485
1975	218	-	32,286	31,444	28,253
1985	228	-	36,470	32,681	28,747
1995	238	-	40,961	32,993	29,060
2005	248	-	45,768	32,353	29,256
2015	258	-	50,899	30,816	29,378
2025	268	-	56,363	28,510	29,453
2035	278	-	62,168	25,621	29,500
2045	288	-	68,324	22,364	29,529
∞	∞	-	∞	$-\infty$	29,575

number of species described by 1934; the number of species names listed in the first part alone is ignored because the second part (Scheerpeltz 1933-1934) includes many species names which had been overlooked in the first part (Bernhauer *et al.* 1910-1926). It is unfortunate that we were not able to discover any estimates for the time period between 1883 and 1934 and that many of the estimates were given in the form "more than" (>) rather than as a more precise figure.

Estimates were made by the same 3 methods as used for Cicindelidae of America north of Mexico. Despite the absurdity of the early (1787-1806) estimates made by the cubic method, the cubic estimates do provide the best least-sum-of-squares fit to the data, followed by the log quadratic estimates, followed by the logistic estimates. However, the estimates made by the cubic method increase to infinity into the future, the quadratic estimates increase to just over 33,000 (against the year 1994) then decrease to minus infinity, while the estimated upper asymptote by the logistic method is 29,575 ± 25. Judging solely by the expected total for America north of Mexico and the ratio of known to expected species for that region, and in the belief that the proportion of undescribed species for the world is likely to be considerably greater than that for America north of Mexico, we cannot accept the estimates made by the logistic method and have already explained reasons for rejection of the cubic and log quadratic methods. The ever-increasing slope produced by the cubic method indicates that there has not been sufficient reduction in species descriptions in recent years to cause an upper levelling off of the line calculated by that method. In brief, we have insufficient data to produce a valid estimate of the world total of species of Staphylinidae by an acceptable method and we have shown that the use of trend curves for this purpose is simplistic because of the nature of the data.

Summary

Even when trend line analysis is performed by correct statistical procedures, it is a poor method for estimation of the number of species existing within a taxon. This is because it attempts to relate the number of species described to time, and involves several implicit assumptions about the form of the relationship. The assumptions may not be justifiable and are impossible to test.

More direct methods of making estimates are greatly to be preferred. A simple method of making an estimate of the number of species of Staphylinidae of America north of Mexico is described. The result of this estimate (>3,416 species) is contrasted with an estimate made by use of trend lines.

Data are yet inadequate for estimating the number of species of Staphylinidae of the world. Trend line analysis produces unacceptable estimates.

References Cited

Arnett, R. H. 1961. The beetles of the United States (a manual for identification). Washington; Catholic Univ. of America Press, pt. 2, p. 211-368.

ARNETT, R. H. 1967. Present and future systematics of the Coleoptera in North America. Ann. Ent. Soc. Am. 60:162-170.

BERNHAUER, M., SCHEERPELTZ, O. & SCHUBERT, K. (*in*) SCHENKLING, S. (ed.) 1910-1926. Coleopterorum catalogus. Berlin; W. Junk, vol. 5, partes 19, 29, 40, 57, 67, 82 (Staphylinidae pt. 1, vol. 1-6), p. 1-988.

BLISS, C. I. 1967. Statistics in biology. vol. 1. New York; McGraw-Hill, p. 1-XIV, 1-558.

BLISS, C. I. 1970. Ibid. vol. 2, p. 1-XIII, 1-639.

CAMPBELL, J. M. 1969. A revision of the New World Oxyporinae (Coleoptera: Staphylinidae). Can. Ent. 101:225-268.

CAMPBELL, J. M. 1973a. A revision of the genus *Tachinus* (Coleoptera: Staphylinidae) of North and Central America. Mem. Ent. Soc. Can. 90:1-137.

CAMPBELL, J. M. 1973b. A revision of the genus *Tachinomorphus* (Coleoptera: Staphylinidae) of North and Central America. Can. Ent. 105:1015-1034.

CAMPBELL, J. M. 1975a. New species and records of *Tachinus* (Coleoptera: Staphylinidae) from North America. Can. Ent. 107:87-94.

CAMPBELL, J. M. 1975b. A revision of the genera *Coproporus* and *Cilea* (Coleoptera: Staphylinidae) of America north of Mexico. Can. Ent. 107:175-216.

CAMPBELL, J. M. 1976. A revision of the genus *Sepedophilus* Gistel (Coleoptera: Staphylinidae) of America north of Mexico. Mem. Ent. Soc. Can. 99:1-89.

CROSSKEY, R. W. AND WHITE, G. B. 1977. The Afrotropical Region. A recommended term in zoogeography. J. Nat. Hist. 11:541-544.

DUVIVIER, A. 1883. Enumération des staphylinides décrits depuis la publication du catalogue de MM. Gemminger & de Harold. Ann. Soc. Ent. Belg. 27:91-215.

EDWARDS, J. G. 1949. Coleoptera or beetles east of the Great Plains. Ann Arbor; Edwards, p. 1-181.

ERICHSON, W. F. 1839-40. Genera et species Staphylinorum, Insectorum Coleopterorum familiae. [Berlin]; Morin, p. 1-954, pls. 1-5.

FABRICIUS, J. C. 1775. Systema entomologiae, sistens Insectorum classes, ordines, genera, species, adiectis synonymis, locis, descriptionibus, observationibus. [Flensburg]; Korti, p. 1-30, 1-832.

FABRICIUS, J. C. 1787. Mantissa Insectorum sistens eorum species nuper detectas adiectis characteribus genericis, differentiis specificis, emendationibus, observationibus. [Copenhagen]; Proft, vol. 1, p. I-XX, 1-348.

FABRICIUS, J. C. 1792. Entomologia systematica emendata et aucta secundum classes, ordines, genera, species adjectis synonimis, locis, observationibus, descriptionibus. [Copenhagen]; Proft, vol. 1, p. 1-339, 1-538.

FABRICIUS, J. C. 1798. Supplementum entomologiae systematicae. [Copenhagen]; Proft & Storch, p. 1-572.

FABRICIUS, J. C. 1801. Systema eleutheratorum secundum ordines, genera, species: adiectis synonymis, locis, observationibus, descriptionibus. [Kiel]; Bibliop. Acad. Nov., vol. 2, p. 1-687.

FAGEL, G. 1970. Revision des genres *Procirrus* Latreille, *Palaminus* Erichson, *Oedichirus* Erichson et voisins de la faune africaine (Coleoptera Staphylinidae Paederinae). Ann. Sci. Zool. Mus. Roy. Afr. Centr. Belg. 186:1-444.

FAUVEL, C. A. A. 1872. Faune Gallo-Rhénane ou species des insectes qui habitent la France, la Belgique, la Hollande, le Luxembourg, la Prusse Rhénane, le Nassau et le Valais avec tableaux synoptiques et planches gravées. Caën; Fauvel, Coléoptères, vol. 3, pt. 3, p. 1-214.

FOWLER, W. W. 1888. The Coleoptera of the British Islands. A descriptive account of the families, genera, and species indigenous to Great Britain and Ireland, with notes as to localities, habitats etc. London; Reeve, vol. 2, Staphylinidae, p. 1-444, pl. 1-70.

FRANK, J. H. 1975. A revision of the New World species of the genus *Erichsonius* Fauvel (Coleoptera: Staphylinidae). Coleopt. Bull. 29:177-203.

GANGLBAUER, L. 1895. Die Käfer von Mitteleuropa. Die Käfer der österreichisch-ungarischen Monarchie, Deutschlands, der Schweiz, sowie des französischen und italienischen Alpengebietes. [Vienna]; Carl Gerold's Sohn, vol. 2, p. I-VI, 1-880.

GEMMINGER, M. AND HAROLD, E. VON. 1868. Catalogus Coleopterorum hucusque descriptorum synonymicus et systematicus. [Munich]; Gummi, vol. 2, p. 425-752.

GRAVENHORST, J. L. C. 1806. Monographia Coleopterorum Micropterorum. [Gottingen]; Dieterich, p. I-XVI, 1-236, 1-12, pl. 1.

HAMMOND, P. M. 1975. The phylogeny of a remarkable new genus and species of gymnusine staphylinid (Coleoptera) from the Auckland Islands. J. Ent. (B) 44:153-173.

HERMAN, L. J. 1970. The ecology, phylogeny, and taxonomy of *Stilicolina* (Coleoptera, Staphylinidae, Paederinae). Am. Mus. Novit. 2412:1-26.

HERMAN, L. H. 1972a. A revision of the rove-beetle genus *Charhyphus* (Coleoptera, Staphylinidae, Phloeocharinae). Am. Mus. Novit. 2496:1-16.

HERMAN, L. H. 1972b. Revision of *Bledius* and related genera. Part. 1. The *aequatorialis, mandibularis,* and *semiferrugineus* groups and two new genera (Coleoptera, Staphylinidae, Oxytelinae). Bull. Am. Mus. Nat. Hist. 149:111-254.

HERMAN, L. H. 1975. Revision and phylogeny of the monogeneric subfamily Pseudopsinae for the world (Staphylinidae, Coleoptera). Bull. Am. Mus. Nat. Hist. 155:241-318.

HERMAN, L. H. 1976. Revision of *Bledius* and related genera. Part II. The *armatus, basalis,* and *melanocephalus* groups (Coleoptera, Staphylinidae, Oxytelinae). Bull. Am. Mus. Nat. Hist. 157:171-172.

HERMAN, L. H. 1977. Revision and phylogeny of *Zalobius, Asemobius,* and *Nanobius,* new genus (Coleoptera, Staphylinidae, Piestinae). Bull. Am. Mus. Nat. Hist. 159:45-86.

HOEBEKE, E. R. 1976. A revision of the genus *Xenodusa* (Staphylinidae, Aleocharinae) for North America. Sociobiology 2:108-143.

KISTNER, D. H. 1976. Revision and reclassification of the genus *Goniusa* Casey with a larval description and ant host records (Coleoptera: Staphylinidae). Sociobiology 2:83-95.

LINNÉ, C. VON. 1758. Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. [Stockholm]; Salvi, 10th ed^{n}., vol. 1, p. 1-823.

MANNERHEIM, C. G. VON. 1831. Précis d'un nouvel arrangement de la famille des brachélytres, de l'ordre des insectes coléoptères. Mem. Acad. Imp. Sci. St.-Petersbourg 1:415-501.

SCHEERPELTZ, O. (*in*) SCHENKLING, S. (ed.) 1933-1934. Coleopterorum catalogus. Berlin; W. Junk, vol. 6, partes 129, 130 (Staphylinidae pt. 2, vol. 7-8, addenda et corrigenda), p. 989-1881.

SEEVERS, C. H. 1957. A monograph on the termitophilous Staphylinidae (Coleoptera). Fieldiana: Zoology 40:1-334.

SEEVERS, C. H. 1965. The systematics, evolution and zoogeography of staphylinid beetles associated with army ants (Coleoptera, Staphylinidae). Fieldiana: Zoology 47:137-351.

SMETANA, A. 1971a. Revision of the tribe Quediini of America north of Mex-

ico (Coleoptera: Staphylinidae). Mem. Ent. Soc. Can. 79: I-VI + 1-303.
SMETANA, A. 1971b. Ibid. Supplementum. Can. Ent. 103:1833-1848.
SMETANA, A. 1973. Ibid. Supplementum 2. Can. Ent. 105:1421-1434.
SMETANA, A. 1976. Ibid. Supplementum 3. Can. Ent. 108:169-184.
STEYSKAL, G. C. 1965. Trend curves of the rate of species description in zoology. Science 149:880-882.
ULRICH, W. G. 1975. Monographie der Gattung *Tachinus* Gravenhorst (Coleoptera: Staphylinidae), mit Bemerkungen zur Phylogenie und Verbreitung der Arten. Doctoral dissertation; Kiel; Christian-Albrechts-Universität, p. 1-365. pls. 1-61.
ULRICH, W. G. AND CAMPBELL, J. M. 1974. A revision of the apterus-group of the genus *Tachinus* Gravenhorst (Coleoptera: Staphylinidae). Can. Ent. 106:627-644.
WHITE, R. E. 1975. Trend curves of the rate of species description for certain North American Coleoptera. Coleopt. Bull. 29:281-295.

BOOK REVIEW

Beetles from the early Russian explorations of the West Coast of North America 1815-1857, ed. by E. Gorton Linsley. 1978. Reprint edition, by Arno Press Inc., Three Park Avenue, New York, NY 10016. Hardbound, ca. 540 p., $40.00.

When a 5-cent cup of coffee costs 35 cents, a 15-cent beer costs 75 cents, and a 3-cent letter costs 15 cents, it is neither surprising nor particularly obscene that a 10-dollar book costs 40 dollars. The question we must ask is, "Is this indeed a 10-dollar book?" There obviously is a market for reprint editions of important but scarce publications; are we the market?

This reprint edition includes a brief note by Keir B. Sterling about the collectors and students of materials secured in Imperial Russian enclaves in western North America in the early 1800's, plus 8 *alpha*-taxonomy articles about beetles published between 1840 and 1860: Mannerheim (6), Ménétriés (1), Motschulsky (1)—a bit over 500 pages reprinted from mostly Russian journals, variously in French, Latin, or German. This is neither more nor less than a bound collection of reprints, neither freshly edited nor consecutively paginated.

In the sense of cost of preparation, quality of reproduction, news to science, and the like, this definitely is not a 10-dollar book. But, that it is not coffee-table quality is very much beside the point.

I judge that this certainly is a 10-dollar book—one that will find a comfortable niche on my shelf and be consulted from time to time—for these reasons: The selection of material is such that access is enhanced; the papers are an important historical resource for beetle taxonomists; and the original papers are not otherwise readily available to most workers. However, I can see no use for it to other than practicing taxonomists.

You will ask me if *I* would pay 40 dollars. Well . . . that's a lot of 75-cent beers . . .

—D.R.W.